LAB -04(Test for equality of two population proportion)

Moushreeta Debroy(22122132)

2022-11-08

AIM:Our objective is to test the percentage of male patient who smokes and percentage of female patient who smokes are equal or not.

ABOUT THE DATASET:Here we are working with blood pressure dataset.It consists of columns “patient number”,“blood pressure abnormality”,“level of hemoglobin”,“sex”,“smoking”,“age” etc.we found this dataset from kaggle.

INTRODUCTION: TEST OF PROPORTION;A test of proportion will assess whether or not a sample from a population represents the true proportion from the entire population. Z TEST:A Z-test is any statistical test for which the distribution of the test statistic under the null hypothesis can be approximated by a normal distribution. Z-tests test the mean of a distribution. Here we are using Z test as the sample size is greater than 30.

```
data<-read.csv('data.csv')

View(data)

head(data)
```

```
##   Patient_Number Blood_Pressure_Abnormality Level_of_Hemoglobin
## 1              1                          1               11.28
## 2              2                          0                9.75
## 3              3                          1               10.79
## 4              4                          0               11.00
## 5              5                          1               14.17
## 6              6                          0               11.64
##   Genetic_Pedigree_Coefficient Age BMI Sex Pregnancy Smoking
Physical_activity
## 1                         0.90  34  23   1         1       0
45961
## 2                         0.23  54  33   1        NA       0
26106
## 3                         0.91  70  49   0        NA       0
9995
## 4                         0.43  71  50   0        NA       0
10635
## 5                         0.83  52  19   0        NA       0
15619
## 6                         0.54  23  48   0        NA       1
27042
##   salt_content_in_the_diet alcohol_consumption_per_day Level_of_Stress
## 1                    48071                          NA               2
```

```
## 2                    25333                            205                3
## 3                    29465                             67                2
## 4                     7439                            242                1
## 5                    49644                            397                2
## 6                     7513                             NA                3
##   Chronic_kidney_disease Adrenal_and_thyroid_disorders
## 1                      1                             1
## 2                      0                             0
## 3                      1                             0
## 4                      1                             0
## 5                      0                             0
## 6                      0                             0

set.seed(200)
```

#sample 1

```
Male<-subset(data,Sex==0)
```

#sample 2

```
Female<-subset(data,Sex==1)

library(dplyr)

## Warning: package 'dplyr' was built under R version 4.2.2

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union

count(Male)#sample size 1

##      n
## 1 1008

count(Female)#sample size 2

##     n
## 1 992

count(Male,Smoking==1)#number of male smokers

##   Smoking == 1   n
## 1        FALSE 494
## 2         TRUE 514
```

```
count(Female,Smoking==1)#number of female smokers
```

```
##   Smoking == 1   n
## 1        FALSE 487
## 2         TRUE 505
```

P1:Proportion of male patient who smokes

P2:Proportion of female patient who smokes H0::P1=P2 H1:P1!=P2

```
x=c(514,505)
n=c(1008,992)
prop.test(x,n,alternative="two.sided",conf.level = 0.95)
```

```
##
##  2-sample test for equality of proportions with continuity correction
##
## data:  x out of n
## X-squared = 2.0729e-29, df = 1, p-value = 1
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  -0.04381961  0.04551572
## sample estimates:
##    prop 1    prop 2
## 0.5099206 0.5090726
```

Conclusion:Here p value is greater than alpha(level of significance).So,here we accept the null hypothesis.so,proportion of male patient who smokes and proportion of female patient who smokes are equal.